

# Using baseline-dependent window functions for data compression and field-of-interest shaping in radio interferometry

M. T. Atemkeng<sup>1\*</sup>, O. M. Smirnov<sup>12</sup>, C. Tasse<sup>31</sup>, G. Foster<sup>12</sup>, J. Jonas<sup>12</sup>

<sup>1</sup>*Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown, 6140, South Africa*

<sup>2</sup>*SKA South Africa, 3rd Floor, The Park, Park Road, Pinelands, 7405, South Africa*

<sup>3</sup>*GEPI, Observatoire de Paris, CNRS, Universite Paris Diderot, 5 place Jules Janssen, 92190 Meudon, France*

Accepted 2016 July 07. Received 2016 June 09; in original form 2016 March 09.

## ABSTRACT

In radio interferometry, observed visibilities are intrinsically sampled at some interval in time and frequency. Modern interferometers are capable of producing data at very high time and frequency resolution; practical limits on storage and computation costs require that some form of data compression be imposed. The traditional form of compression is a simple averaging of the visibilities over coarser time and frequency bins. This has an undesired side effect: the resulting averaged visibilities “decorrelate”, and do so differently depending on the baseline length and averaging interval. This translates into a non-trivial signature in the image domain known as “smearing”, which manifests itself as an attenuation in amplitude towards off-centre sources. With the increasing fields of view and/or longer baselines employed in modern and future instruments, the trade-off between data rate and smearing becomes increasingly unfavourable. In this work we investigate alternative approaches to low-loss data compression. We show that averaging of the visibility data can be treated as a form of convolution by a boxcar-like window function, and that by employing alternative baseline-dependent window functions a more optimal interferometer smearing response may be induced. In particular, we show improved amplitude response over a chosen field of interest, and better attenuation of sources outside the field of interest. The main cost of this technique is a reduction in nominal sensitivity; we investigate the smearing vs. sensitivity trade-off, and show that in certain regimes a favourable compromise can be achieved. We show the application of this technique to simulated data from the Karl G. Jansky Very Large Array (VLA) and the European Very-long-baseline interferometry Network (EVN).

**Key words:** Instrumentation: interferometers, Methods: data analysis, Methods: numerical, Techniques: interferometric

## 1 INTRODUCTION

A radio interferometer measures complex quantities called *visibilities*, which, following the van Cittert-Zernike relation (Thompson 1999; Thompson et al. 2001), correspond to Fourier modes of the sky brightness distribution, corrupted by various instrumental and atmospheric effects. One particular effect, known as *time* and *bandwidth decorrelation* (or its equivalent in the image plane referred as smearing) occurs when the visibilities are averaged over a time and frequency bin of non-zero extent (Bridle & Schwab 1989, 1999).

This unavoidably happens in the correlator (since the correlator output is, by definition, an average measurement over some interval), and also if data is further averaged in post-correlation for the purposes of compression and to reduce computational cost.

The effect of smearing is mainly a decrease in the amplitude of off-axis sources. This is easy to understand: the visibility contribution of a point source of flux  $S$  located in the direction given by the unit vector  $\boldsymbol{\sigma}$  is given by

$$V = S \exp \left\{ \frac{2\pi i}{\lambda} \mathbf{u} \cdot (\boldsymbol{\sigma} - \boldsymbol{\sigma}_0) \right\}, \quad (1)$$

where  $\mathbf{u}$  is the baseline vector (in a coordinate system fixed to the sky), and  $\boldsymbol{\sigma}_0$  is the phase centre (or fringe stopping

\* E-mail: m.atemkeng@gmail.com

centre) of the observation. The complex phase term above rotates as a function of frequency (due to the inverse scaling with  $\lambda$ ) and time (due to the fact that  $\mathbf{u}$  changes with time, at least in an Earth- or orbit-based interferometer). Taking a rotating complex vector average over a time/frequency bin then results in a net loss of amplitude. The effect increases with baseline length and distance from phase centre. Besides reducing apparent source flux, smearing also distorts the point spread function (PSF), since different baselines (and thus different Fourier modes) are attenuated differently. The issue of time-frequency averaging has been addressed in the past by among others Thompson et al. (2008), where a Gaussian taper has been used to eliminate smearing at the edges of the field of view (FoV). However, the problem of eliminating smearing to about 1% or less within the FoV while compressing the data to an acceptable level has not been satisfactorily addressed before.

Throughout this work we use the term Field of Interest (FoI) which we differentiate from the FoV. For this work the FoV is related to angular scale of the primary beam where as the FoI is a parameter of the scientific observation case, which may be related to the size of the primary beam but is not a necessary requirement. Thus, the FoI can be an adjustable parameter in the window functions we present.

In the era of large interferometers, where computation (and thus data size) becomes one of the main cost drivers, it is in principle desirable to average the data down as much as possible, without compromising the science goals. There are natural limits to this: firstly, we still need to critically sample the  $uv$ -plane, secondly, we need to retain sufficient spectral resolution, thirdly, we do not want to average (at least pre-calibration) beyond the natural variation of the calibration parameters, and fourthly, we want to keep smearing at acceptable levels in order not to lose too much signal. In this work, we concentrate specifically on the decorrelation/smearing problem. Typical observation cases are:

- Surveys where it is desirable to have a flat response across the largest possible FoI. This can be achieved by facet imaging but at a computational cost. By increasing the size of the facets a better balance between computational cost and image response can be achieved.
- Deep imaging where it is useful to suppress bright sources in the primary beam sidelobes outside of the FoI. Even with smearing of sources far from the phase centre bright sources can contribute significant flux and PSF side-lobe artefacts to the image.
- Very Long Baseline Interferometry (VLBI), decorrelation is more severe, so the effective FoI is determined by the smallest time/frequency bin size that a correlator can support, and is normally much smaller than the primary beam (Keimpema et al. 2015). Modern VLBI correlators overcome this by employing a technique where the signal is correlated relative to multiple phase centres simultaneously, thus effectively “tiling” the primary beam by multiple FoIs. This has a computational cost that scales linearly with the number of phase centres.

Smearing can be seen to be a useful side effect, anything outside the desired FoI (by definition) is unwanted signal. The primary beam pattern of any real antenna features sidelobes and backlobes that extend across the entire sky, albeit at a relatively faint level. The faintness makes sidelobes use-

less for imaging any but the brightest sources. However, the sum total signal from all the sources in the primary beam sidelobes, modulated by their PSF sidelobes, contributes an unwanted global background called the *far sidelobe confusion noise* (FSCN), in very deep observations this may in principle become a bottleneck (Smirnov et al. 2012). In other cases, individual extremely bright radio sources such as Cygnus A or Cassiopeia A can contribute confusing signal from even the most distant sidelobe: the LOFAR telescope (van Haarlem et al. 2013) has to deal with these so-called “A-team” sources on a routine basis. By suppressing distant off-axis sources, smearing somewhat alleviates both the FSCN and A-team problems.

When considering a short sequence of visibilities measured on one baseline, we can consider averaging as a convolution of the true visibility by a boxcar function corresponding to the  $uv$ -extent of the averaging bin, followed by sampling at the centre of each bin. Convolution in the visibility plane corresponds to multiplication of the image by an *image plane response function* that is the Fourier transform of the convolution kernel i.e the window function; the Fourier transform of a boxcar is a sinc-type taper.

If we consider the entire  $uv$ -plane, averaging is only a pseudo-convolution, since the different  $uv$ -bins (and thus their boxcars) will have different sizes and shapes as determined by baseline length and orientation. Still, we can qualitatively view smearing as some kind of cumulative effect of an ensemble of image-plane tapers corresponding to all the different boxcars<sup>1</sup>.

What if we were to employ weighted averaging instead of simple averaging (whether in the correlator, or in post-processing)? This would correspond to a pseudo-convolution of the  $uv$ -plane by some ensemble of *window functions*, different from boxcars, which would obviously yield different image-plane tapers, and thus result in different smearing response. Filter theory suggests that a window function can be tuned to achieve some desired tapering response. An optimal taper would be one that was maximal across the desired FoI, and minimal outside it. In this work, we apply filter theory to derive a set of baseline-dependent window functions (BDWFs)<sup>2</sup> that approximate this more optimal smearing behaviour. The trade-off is an increase in thermal noise, since minimum noise can only be achieved with unweighted averaging. We show that this effect can be partially mitigated through the use of *overlapping window functions*. Offringa et al. (2012) have investigated a similar approach

<sup>1</sup> For completeness, we should note that this “smearing taper” is not the only tapering effect at work in interferometric imaging. Firstly, antennas have a non-zero physical extent: a measured visibility is already convolved by the aperture illumination functions of each pair of antennas. The resulting image-plane taper is exactly what the primary beam is. Secondly, most imaging software employs convolutional gridding followed by a fast Fourier transform, which produces an additional taper that suppresses aliasing of sources from outside the imaged region.

<sup>2</sup> BDWFs are functions of a baseline’s length and orientation; the East-West component rotates as a function of time, while the South-North does not. For the same length of time, two equal length baselines with different orientations will sweep unequal  $uv$ -space and therefore result in a different degree of decorrelation.

in the context of suppressing signals towards specific off-axis sources.

In the era of the Square Kilometre Array (SKA) and its pathfinders, where dealing with the huge data volumes is one of the main challenges, use of BDWFs potentially offers additional leverage in optimising radio observations. Decreased smearing across the FoI allows for more aggressive data averaging, thus reducing storage and compute costs. The trade-off is a loss of sensitivity, which pushes up observational time requirements. However, the decrease in smearing and noise from A-team sources could, conceivably, make up for some of the nominal sensitivity loss. In the VLBI case, use of BDWFs potentially offers an increase in effective FoI at a given correlator dump rate, or equivalently, the ability to tile the primary beam with fewer phase centres, allowing for smaller correlators.

## 2 OVERVIEW AND PROBLEM STATEMENT

The following formalism deals with visibilities both as functions (i.e. entire distributions on the  $uv$ -plane), and single visibilities (i.e. values of those functions at a specific point). To avoid confusion between functions in functional notation and their values, we will use  $\mathcal{V}$  or  $\mathcal{V}(u, v)$  to denote functions, and  $V$  to denote individual visibilities. Likewise,  $\mathcal{I}(l, m)$  denotes a function on the  $lm$ -plane i.e. an image. The symbol  $\delta$  always denotes the Kronecker delta-function.

Depending on whether we want to consider polarisation or not,  $\mathcal{V}$  can be taken to represent either scalar (complex) visibilities, or  $2 \times 2$  complex visibility matrices as per the radio interferometer measurement equation (RIME) formalism (Smirnov 2011). Likewise,  $\mathcal{I}$  can be treated as a scalar (total intensity) image, or a  $2 \times 2$  brightness matrix distribution. The derivations below are valid in either case.

We shall use the symbols  $\mathbf{u} = (u, v)$  or  $\mathbf{u} = (u, v, w)$  to represent baseline coordinates in units of wavelength.

### 2.1 Visibility and relation with the sky

An interferometer array measures the quantity  $\mathcal{V}(u, v, w)$ , known as the visibility function. Here, the coordinates  $u, v$  and  $w$  are vector components in units of wavelength, describing the distance between antennas  $p$  and  $q$ , called the *baseline*. The  $w$  axis is oriented towards the *phase centre* of the observed field, while  $u$  points east and  $v$  north. Given a sky distribution  $\mathcal{I}_0(l, m)$ , where  $l, m$  are the direction cosines, the nominal observed visibility is given by the van Cittert-Zernike theorem (Thompson 1999; Thompson et al. 2001) as

$$\mathcal{V}^{\text{nom}}(u, v) = \iint_{lm} \frac{\mathcal{I}_0(l, m)}{\sqrt{1-l^2-m^2}} e^{-2\pi i \phi(u, v, w)} dldm, \quad (2)$$

where  $\phi(u, v, w) = ul + vm + w(n-1)$ , and  $n = \sqrt{1-l^2-m^2}$  (the  $n-1$  term comes about when fringe stopping is in effect, i.e. when the correlator introduces a compensating delay to ensure  $\phi = 0$  at the centre of the field, otherwise the term is simply  $n$ ).

Given a pair of antennas  $p$  and  $q$  forming a baseline  $\mathbf{u}_{pq} = (u_{pq}, v_{pq}, w_{pq})$ , and taking into account the *primary*

*beam* patterns  $\mathcal{E}_p(l, m)$  and  $\mathcal{E}_q(l, m)$  that define the directional sensitivity of the antennas, this becomes

$$\mathcal{V}_{pq}(u, v) = \iint_{lm} \frac{\mathcal{E}_p \mathcal{I}_0 \mathcal{E}_q^H}{\sqrt{1-l^2-m^2}} e^{-2\pi i \phi(u, v, w)} dldm, \quad (3)$$

where  $^H$  represents the conjugate transpose. The first term being integrated is the *apparent sky seen by baseline*  $pq$ ,

$$\mathcal{I}_{pq} = \frac{\mathcal{E}_p \mathcal{I}_0 \mathcal{E}_q^H}{\sqrt{1-l^2-m^2}}, \quad (4)$$

which in general can be variable in time and frequency. For simplicity, let us assume that both the sky and the primary beam are constant (invariant in time and frequency), and that the primary beam is the same for all stations. All baselines will then see the same apparent sky throughout the measurement process. Let us designate this by  $\mathcal{I}$ . Assuming a small FoI ( $n \rightarrow 1$ ) and/or a co-planar array ( $w = 0$ ), the above equation becomes a simple 2D Fourier transform :

$$\mathcal{V}(u, v) = \iint_{lm} \mathcal{I} e^{-2\pi i (ul + vm)} dldm, \quad (5)$$

or in functional form,

$$\mathcal{V} = \mathcal{F}\{\mathcal{I}\}, \quad \mathcal{I} = \mathcal{F}^{-1}\{\mathcal{V}\}. \quad (6)$$

we will refer to  $\mathcal{V}$  as the *ideal* visibility distribution (as opposed to the *measured* distribution, which is corrupted by averaging in the correlator, as we will explore below).

Note that the effect of the primary beam can alternatively be expressed in terms of a convolution with its Fourier transform, the *aperture illumination function*  $\mathcal{A}_p(u, v)$ . In functional form:

$$\mathcal{V}_{pq} = \mathcal{A}_p \circ \mathcal{V}_{pq}^{\text{nom}} \circ \mathcal{A}_q^H. \quad (7)$$

### 2.2 Imaging, averaging and convolution

Earth rotation causes the baseline to rotate in time, the baseline in units of wavelength can be treated as a function of frequency and time (from this point on-wards we shall assume that the sky is constant across the range of frequencies being observed):

$$\mathbf{u}_{pq}(t, \nu) = \mathbf{u}_{pq}(t) \nu / c. \quad (8)$$

This, in turn, allows us to rewrite the visibility in eq. (5) as a per-baseline function of  $t, \nu$ :

$$\mathcal{V}_{pq}(t, \nu) = \iint_{lm} \mathcal{I} e^{-2\pi i (u_{pq}(t, \nu)l + v_{pq}(t, \nu)m)} dldm. \quad (9)$$

Synthesis imaging recovers the so-called “dirty image”: the inverse Fourier transform of the measured visibility distribution  $\hat{\mathcal{V}}$  sampled by a number of baselines  $pq$  at discrete time/frequency points. Inverting the Fourier transform produces the dirty image:

$$\mathcal{I}^D = \mathcal{F}^{-1}\{\mathcal{W} \cdot \hat{\mathcal{V}}\}, \quad (10)$$

where  $\mathcal{W}$  is the (weighted) sampling function – a “bed-of-nails” function that is non-zero at points where we are sampling a visibility, and zero elsewhere. If  $\hat{\mathcal{V}} = \mathcal{V}$ , then this can also be expressed as a convolution of the apparent sky by the *point spread function*  $\mathcal{P}$ :

$$\mathcal{I}^D = \mathcal{P} \circ \mathcal{I}, \quad \mathcal{P} = \mathcal{F}^{-1}\{\mathcal{W}\}. \quad (11)$$

Designating each baseline as  $pq$ , and each time/frequency point as  $t_k, \nu_l$ , we can represent  $\mathcal{W}$  by a sum of “single-nail” functions  $\mathcal{W}_{pqkl}$ :

$$\mathcal{W} = \sum_{pqkl} \mathcal{W}_{pqkl} = \sum_{pqkl} W_{pqkl} \delta_{pqkl}, \quad (12)$$

where  $\delta_{pqkl}$  is a delta-function shifted to the  $uv$ -point being sampled:

$$\delta_{pqkl}(\mathbf{u}) = \delta(\mathbf{u} - \mathbf{u}_{pq}(t_k, \nu_l)) \quad (13)$$

and  $W_{pqkl}$  is the associated weight. The Fourier transform being linear, we can rewrite eq. (10) as

$$\mathcal{I}^D = \sum_{pqkl} W_{pqkl} \mathcal{F}^H \{\hat{\mathcal{V}}_{pqkl}\}, \quad (14)$$

where

$$\hat{\mathcal{V}}_{pqkl} = \delta_{pqkl} \hat{\mathcal{V}}_{pq}(t_k, \nu_l) \quad (15)$$

i.e. the visibility distribution corresponding to the single visibility sample  $pqkl$ . We can further rewrite eq. (10) again as

$$\mathcal{I}^D = \sum_{pqkl} W_{pqkl} \mathcal{F}^{-1} \{\hat{\mathcal{V}}_{pqkl}\}, \quad (16)$$

which shows that the dirty image  $\mathcal{I}^D$  can be seen as a weighted sum of images corresponding to the individual visibility samples  $pqkl$  (each such image essentially being a single fringe pattern).

In the ideal case, we would be measuring instantaneous visibility samples, and (assuming no other instrumental corruptions), we would have  $\hat{\mathcal{V}} \equiv \mathcal{V}$ , with

$$\hat{\mathcal{V}}_{pq}(t_k, \nu_l) = \mathcal{V}(\mathbf{u}_{pq}(t_k, \nu_l)), \quad (17)$$

and consequently,

$$\hat{\mathcal{V}}_{pqkl} = \delta_{pqkl} \mathcal{V}, \quad (18)$$

resulting in what we will call the *ideal* dirty image  $\mathcal{I}^{\text{DI}}$ :

$$\mathcal{I}^{\text{DI}} = \sum_{pqkl} W_{pqkl} \mathcal{P}_{pqkl} \circ \mathcal{I}, \quad \mathcal{P}_{pqkl} = \mathcal{F}^{-1} \{\delta_{pqkl}\} \quad (19)$$

That is, in the ideal case, each term in the weighted sum is equal to the apparent sky  $\mathcal{I}$  convolved with a PSF representing a single visibility sample,  $\mathcal{P}_{pqkl}$ .

However, an actual interferometer is necessarily non-ideal, in that it can only measure the average visibility over a some time-frequency bin given by the *time* and *frequency sampling intervals*  $\Delta t, \Delta \nu$ , which we will call the *sampling bin*

$$\mathbf{B}_{kl}^{[\Delta t \Delta \nu]} = \left[ t_k - \frac{\Delta t}{2}, t_k + \frac{\Delta t}{2} \right] \times \left[ \nu_l - \frac{\Delta \nu}{2}, \nu_l + \frac{\Delta \nu}{2} \right], \quad (20)$$

This measurement can be represented by an integration:

$$\hat{\mathcal{V}}_{pqkl} = \frac{1}{\Delta t \Delta \nu} \iint_{\mathbf{B}_{kl}^{[\Delta t \Delta \nu]}} \mathcal{V}(\mathbf{u}_{pq}(t, \nu)) d\nu dt. \quad (21)$$

Inverting the relation of eq. (8), we can change variables to express this as an integration over the corresponding bin  $\mathbf{B}_{pqkl}^{[uv]}$  in  $uv$ -space:

$$\hat{\mathcal{V}}_{pqkl} = \frac{1}{\Delta t \Delta \nu} \iint_{\mathbf{B}_{pqkl}^{[uv]}} \mathcal{V}_{pq}(u, v) \left| \frac{\partial(t, \nu)}{\partial(u, v)} \right| du dv, \quad (22)$$

where  $\mathbf{B}_{pqkl}^{[uv]}$  is the corresponding bin in  $uv$ -space. Note that the sampling bins in  $tv$ -space are perfectly rectangular (Fig. 1, right) and do not depend on baseline (assuming baseline-independent averaging), while the sampling bins in  $uv$ -space are elliptical arcs, and do depend on baseline (hence the extra  $pq$  index). Assuming a bin small enough that the fringe rate  $\partial \mathbf{u} / \partial t$  is approximately constant over the bin, we then have

$$\hat{\mathcal{V}}_{pqkl} \sim \iint_{\mathbf{B}_{pqkl}^{[uv]}} \mathcal{V}(\mathbf{u}) d\mathbf{u}, \quad (23)$$

Now, let us introduce a *normalised boxcar window function*,  $\Pi^{[t\nu]}$

$$\Pi^{[t\nu]}(t, \nu) = \begin{cases} \frac{1}{\Delta t \Delta \nu}, & |t| \leq \Delta t/2, \quad |\nu| \leq \Delta \nu/2 \\ 0, & \text{otherwise,} \end{cases} \quad (24)$$

using which we may re-write eq. (21) as

$$\hat{\mathcal{V}}_{pqkl} = \iint_{\infty} \mathcal{V}_{pq}(t, \nu) \Pi^{[t\nu]}(t - t_k, \nu - \nu_l) dt d\nu, \quad (25)$$

which can also be expressed as a convolution:

$$\hat{\mathcal{V}}_{pqkl} = [\mathcal{V}_{pq} \circ \Pi^{[t\nu]}](t_k, \nu_l), \quad (26)$$

Likewise, eq. (22) can also be rewritten as a convolution in  $uv$ -space:

$$\hat{\mathcal{V}}_{pqkl} = [\mathcal{V}_{pq} \circ \Pi_{pqkl}^{[uv]}](\mathbf{u}_{pq}(t_k, \nu_l)), \quad (27)$$

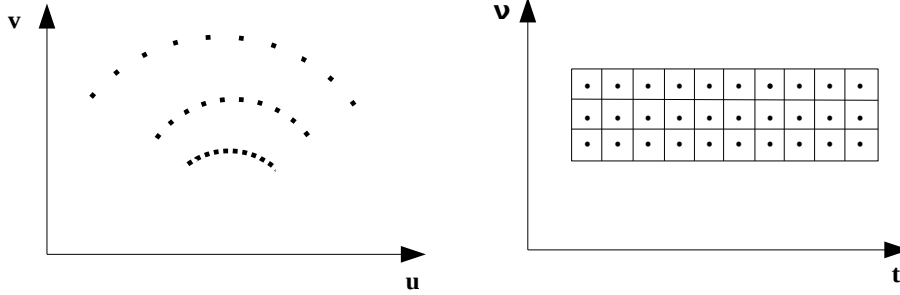
where  $\Pi_{pqkl}^{[uv]}$  is a boxcar-like window function that corresponds to bin  $\mathbf{B}_{pqkl}^{[uv]}$  in  $uv$ -space (and also includes the determinant term of eq. 22). This makes it explicit that each averaged visibility is drawn from a convolution of the underlying visibilities with a boxcar-like window function.

Note what eq. (27) does and does not say. It does say that each individual averaged visibility corresponds to convolving the true visibilities by some window function. However, this window function is different for each baseline  $pq$  and time/frequency sample  $t_k, \nu_l$  (which is emphasised by the subscripts to  $\Pi^{[uv]}$  in the equations above). Averaging is thus not a “true” convolution, since the convolution kernel changes at every point in the  $uv$ -plane. We will call this process a *pseudo-convolution*, and the kernel being convolved with ( $\Pi_{pqkl}^{[uv]}$ ) an example of a *baseline-dependent window function* (BDWF). In subsequent sections we will explore alternative BDWFs.

In actual fact, a correlator (or an averaging operation in post-processing) deals with averages of discrete and noisy samples, rather than a continuous integration. Ignoring the complexities of correlator implementation, let us cast this process in terms of a simple averaging operation. That is, assume we have a set of *hi-res* or *sampled visibilities* on a high-resolution time/frequency grid  $t_i, \nu_j$ :

$$V_{pqij}^S = \mathcal{V}_{pq}(t_i, \nu_j) + \mathcal{N}[\sigma_{pqij}^{(s)}], \quad (28)$$

where  $\mathcal{V}_{pq}$  is given by eq. (9), and  $\mathcal{N}$  represents the visibility noise term, which is a complex scalar or complex  $2 \times 2$  matrix with the real and imaginary parts being independently drawn from a zero-mean normal distribution with the indicated r.m.s. (Wrobel & Walker 1999). The noise term is not



**Figure 1.** Schematic of  $uv$ -coverage for regularly spaced time-frequency samples (left:  $uv$ -space, right:  $tv$ -space). Baselines with a longer East-West component sweep out longer tracks between successive integration's.

correlated across samples. The *lo-res* or *averaged* or *resampled* visibilities are then a discrete sum:

$$\hat{V}_{pqkl} = \frac{1}{n} \sum_{ij \in B_{kl}} V_{pqij}^S, \quad (29)$$

where  $B_{kl}$  is the set of sample indices  $ij$  corresponding to the *resampling bin*, i.e.

$$B_{kl} = \{ij : t_i \nu_j \in B_{kl}^{[\Delta t \Delta \nu]}\}, \quad (30)$$

and  $n = n_t \times n_\nu$  is the number of samples in the bin. Using the BDWF definitions above, this becomes a conventional discrete convolution (assuming a regular  $tv$  grid):

$$\hat{V}_{pqkl} = \sum_{i,j=-\infty}^{\infty} V_{pqij}^S \Pi^{[t\nu]}(t_i - t_k, \nu_j - \nu_l). \quad (31)$$

In  $uv$ -space, this becomes a discrete convolution on an irregular grid (the  $\mathbf{u}_{ij}$  grid being schematically illustrated by Fig. 1, left):

$$\hat{V}_{pqkl} = \sum_{i,j=-\infty}^{\infty} V_{pqij}^S \Pi_{pqkl}^{[uv]}(\mathbf{u}_{ij} - \mathbf{u}_{kl}), \quad (32)$$

### 2.3 Effect of averaging on the image

In the limit of  $\Delta t, \Delta \nu \rightarrow 0$ , averaging becomes equivalent to sampling. An interferometer must, intrinsically, employ a finitely small averaging interval. The Fourier phase component  $2\pi\phi(u, v, w)$  is a function of frequency and time, with increasing variation over the averaging interval for sources far from the phase centre. The average of a complex quantity with a varying phase then effectively “washes out” amplitude, the effect being especially severe for wide FoIs (for an extensive discussion, see Bregman 2012). In the  $uv$ -plane, this effect is often referred to as *time* and *bandwidth decorrelation*, and *smearing* in the image plane.

The discussion above provides an alternative way to look at decorrelation/smearing. With averaging in effect, the relationship between the measured and the ideal visibility changes to (contrast this to eq. 18):

$$\hat{\mathcal{V}}_{pqkl} = \delta_{pqkl} (\mathcal{V} \circ \Pi_{pqkl}^{[uv]}), \quad (33)$$

Combining this with eq. 16, and using the Fourier convolution theorem, we can see that the dirty image is formed as:

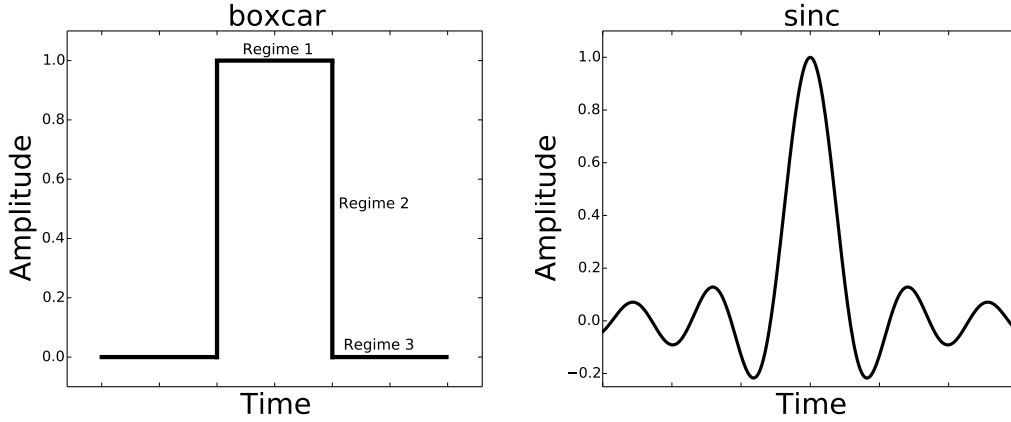
$$\mathcal{I}^D = \sum_{pqkl} W_{pqkl} \mathcal{P}_{pqkl} \circ (\mathcal{I} \cdot \mathcal{T}_{pqkl}), \quad (34)$$

with the apparent sky  $\mathcal{I}$  now tapered by the baseline-dependent *window response function*  $\mathcal{T}_{pqkl}$ , the latter being the inverse Fourier transform of the BDWF:

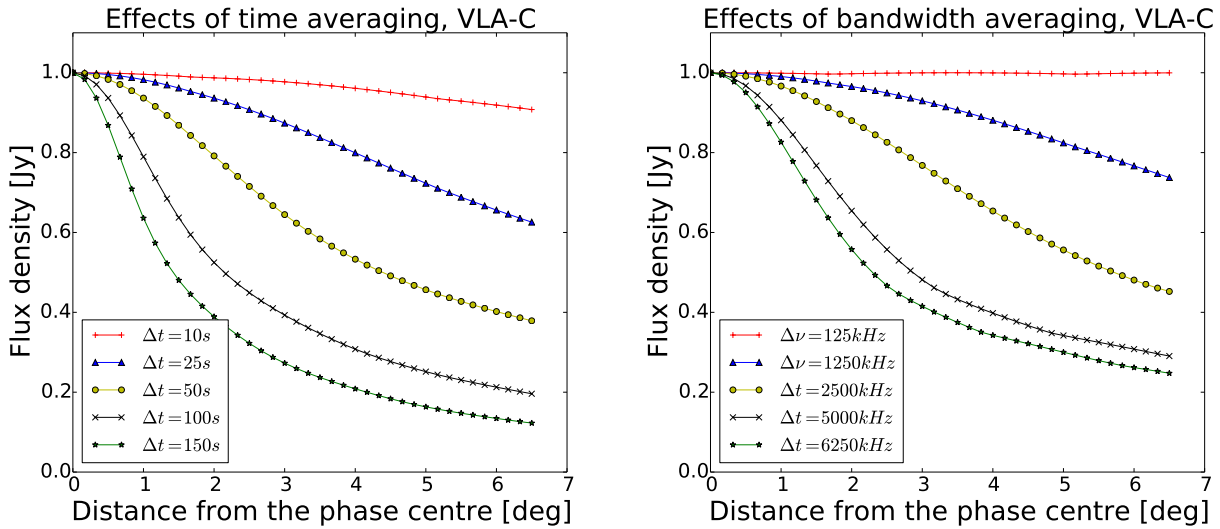
$$\mathcal{T}_{pqkl} = \mathcal{F}^{-1}\{\Pi_{pqkl}^{[uv]}\}. \quad (35)$$

In other words, the dirty image yielded by averaged visibilities (compare this to the ideal dirty image given by eq. 19) is a weighted average of per-visibility dirty images corresponding to a per-visibility tapered sky. The Fourier transform of a boxcar-like function is a sinc-like function, schematically illustrated in 1-D by Fig. 2 (right). Time and bandwidth smearing represents the average effect of all these individual tapers. Shorter baselines correspond to smaller boxcars and wider tapers, longer baselines to larger boxcars and narrower tapers, and are thus more prone to smearing. Fig. 3 (produced by simulating a series of high time-frequency resolution observation using MeqTrees (Noordam & Smirnov 2010), and applying averaging) shows the attenuation of a 1 Jy source as a function of distance from phase centre, for a set of different time and frequency intervals. The simulations correspond to VLA in the C configuration, with an observing frequency of 1.4 GHz. At this frequency, the first null of the primary beam is at  $r \approx 36'$ , and the half-power point is at  $\sim 16'$ , thus we can consider the “conventional” FoI (i.e. the half-power beam width, or HPBW) to be about  $0.5^\circ$  across. Note that the sensitivity of the upgraded VLA, as well as improvements in calibration techniques (Perley 2013), allow imaging to be done in the first primary beam sidelobe as well (and in fact it may be necessary for deep pointing, if only to deconvolve and subtract sidelobe sources), so we could also consider an “extended” FoI out to the second null of the primary beam at  $r \approx 1.25^\circ$ . Whatever definition of the FoI we adopt, Fig. 3 shows that to keep amplitude losses across the FoI to within some acceptable threshold, say 1%, the averaging interval cannot exceed some critical size, say 10 s and 1 MHz. Conversely, if we were to adopt an aggressive averaging strategy for the purposes of data compression, say 50 s and 5 MHz, the curves indicate that we would suffer substantial amplitude loss towards the edge of the FoI.

Finally, note that the curves corresponding to acceptably low values of smearing across the FoI (i.e. up to 25 s and up to 1.25 MHz) have a very gentle slope, with very little suppression of sources *outside* the FoI.



**Figure 2.** Left: boxcar response. In the  $uv$ -plane, this represents the window function corresponding to normal averaging of visibilities. In the image plane, this represents the ideal image-plane response function. Right: sinc response. In the image plane, this represents the window response function corresponding to a boxcar window function in the  $uv$ -plane. In the  $uv$ -plane, this represents the ideal window function.



**Figure 3.** Effects of time and frequency averaging: the apparent intensity of a 1 Jy source, as seen by VLA-C at 1.4 GHz, as a function of distance from phase centre. (Left) Frequency interval fixed at 125 kHz, time interval varies; (right) time interval fixed at 1 s, frequency interval varies.

#### 2.4 The case for alternative BDWFs

The window response or image plane response (IPR) function induced by normal averaging (Fig. 2 (right)) is far from ideal: it either suppresses too much within the FoI, or provides limited benefit to suppressing outside the FoI sources, or both. The optimal image plane response would be a disk-like function, unity within the region bounded by a circle and zero outside this region. In 1-D, it is a boxcar-like, as in Fig. 2 (left).

The BDWF that would produce such a response is sinc-like, as in 1-D presented in Fig. 2 (right). The problem with a sinc is that it has infinite support; applying it over finite-sized bins necessarily means a *truncated* BDWF that results in a sub-optimal taper. The problem of optimal filtering has been well studied in signal processing (usually assuming a

true convolution rather than the pseudo-convolution we deal with here), and we shall apply these lessons below.

The derivations above make it clear that using a different BDWF in place of the conventional boxcar-like  $\Pi^{[uv]}$  could in principle yield a more optimal tapering response. The obvious disadvantage is a loss in sensitivity. Each visibility sample is subject to an independent Gaussian noise term in the real and imaginary part; the noise of the average of a set of samples is minimised when the average is naturally weighted (or unweighted, if the noise is constant across visibilities). Thus, any deviation from a boxcar F must necessarily increase the noise in the visibilities. Below we will study this effect both theoretically and via simulations, to establish whether this trade-off is sensible, and under which conditions.

### 3 APPLYING WINDOW FUNCTIONS TO VISIBILITIES

While visibilities are (usually) regularly sampled in  $tv$ -space, in  $uv^m$ -space this is not so. In frequency, the sampling positions go as  $\sim \lambda^{-1}$ , while in time, baselines with a longer East-West component sweep out longer tracks between successive integrations (Fig. 1). Applying a window function with a constant integration window in  $tv$  space corresponds to different-sized windows in  $uv$ -space. In the case of normal averaging, this results in the boxcar-like window  $\Pi^{[uv]}$  of eq. (27) having a baseline-dependent scale. The scale of the tapering response being inversely proportional to the scale of the window function, this results in more decorrelation (i.e. a narrower achievable FoI) on longer baselines.

By defining our alternative window functions in  $uv$ -space (in units of wavelength), we can attempt to “even out” the decorrelation response across baselines. For a given BDWF  $X(u, v)$ , we have the following recipe for computing resampled visibilities (compare to eq. 32):

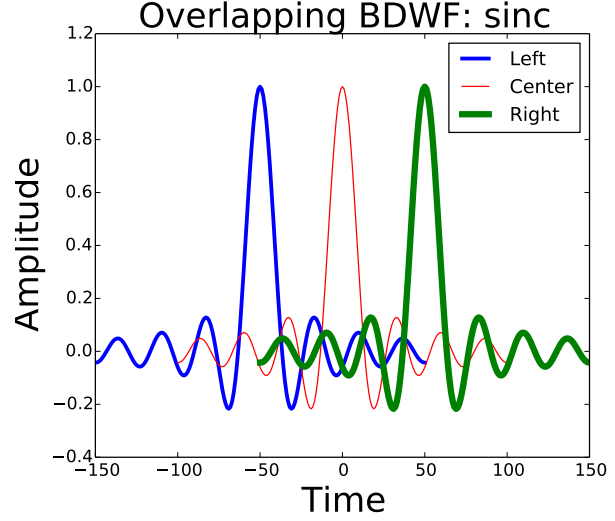
$$\hat{V}_{pqkl} = \frac{\sum_{i,j \in \mathbf{B}_{kl}} V_{pqij}^S X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl})}{\sum_{i,j \in \mathbf{B}_{kl}} X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl})}, \quad (36)$$

where  $\mathbf{u}_{pqkl}$  is the midpoint of the resampling bin  $\mathbf{B}_{kl}$  in  $uv$ -space. The main lobe of the window function then has the same scale across the entire  $uv$ -plane, while the resampling bins have different  $uv$ -sizes. Conversely, in  $tv$ -space the bins are regular, while the main lobe of the effective window function scales inversely with the baseline fringe rate. Furthermore, the window function is truncated at the edge of each bin; on the shortest baselines this truncation is extreme to the point of approaching the boxcar-like  $\Pi^{[uv]}$  (Fig. 4). The downside of this simple approach is twofold. Firstly, while all the window functions above nominally exhibit far lower sidelobes than the boxcar (i.e. more suppression for out-of-FoI sources), they no longer perform particularly well under truncation, with extremely truncated window functions at the shorter baselines becoming boxcar-like. Secondly, taking a weighted sum in eq. 36 increases the noise in comparison to normal averaging.

We use the standard discrete signal processing (DSP) filter terminology to describe the IPR of visibility domain window functions. Window functions or rather their corresponding image-plane response (IPR) can be characterised in terms of various metrics. Some common ones are the peak sidelobe level, the main lobe width and the sidelobes roll-off rate (Smith et al. 1997). In terms of the “ideal” IPR (Fig. 2, left), these correspond to the following desirable traits:

- Maximally conserve the signal within the FoI (“Regime 1” in the figure), and make the transition region (“Regime 2”) as sharp as possible. Both of these correspond to larger main lobe width.
- Attenuate sources outside the FoI (“Regime 3”): this corresponds to a lower peak sidelobe level and higher sidelobes roll-off.

Table A1 (see Appendix A) summarises the performance of different window functions well known in DSP (Smith et al. 1997; Prabhu 2013). This table shows that the IPR of the sinc, the Bessel of the first kind of order zero ( $J_0$ ) (Wat-



**Figure 5.** Overlapping BDWFs representing adjacent resampling bins (right resampling bin, centre resampling bin and left resampling bin). This corresponds to overlap factor  $\alpha = 3$  in time.

son 1995) and all their derivatives with Hamming, Han and Blackman window functions have a larger main lobe, low peak sidelobe level and high sidelobe roll-off compared to others. Hence, they provide the most promising IPRs for our purposes. Note that any window derived from the sinc or Bessel with Hamming, Han or Blackman requires two successive visibility weightings, which may result in amplifying the thermal noise compared to the sinc or Bessel. This makes the sinc and the Bessel more suitable for this work. For this reason, we have chosen to use the sinc and the Bessel window functions to serve as the basis of BDWFs developed in the rest of this paper. We use the following definitions to construct a 2-D sinc and Bessel window functions from their 1-D variants:

$$X(u, v) = X(\kappa u)X(\kappa v), \quad X \equiv \text{sinc} \quad (37)$$

$$X(u, v) = X(\kappa r), \quad r = \sqrt{u^2 + v^2}, \quad X \equiv J_0. \quad (38)$$

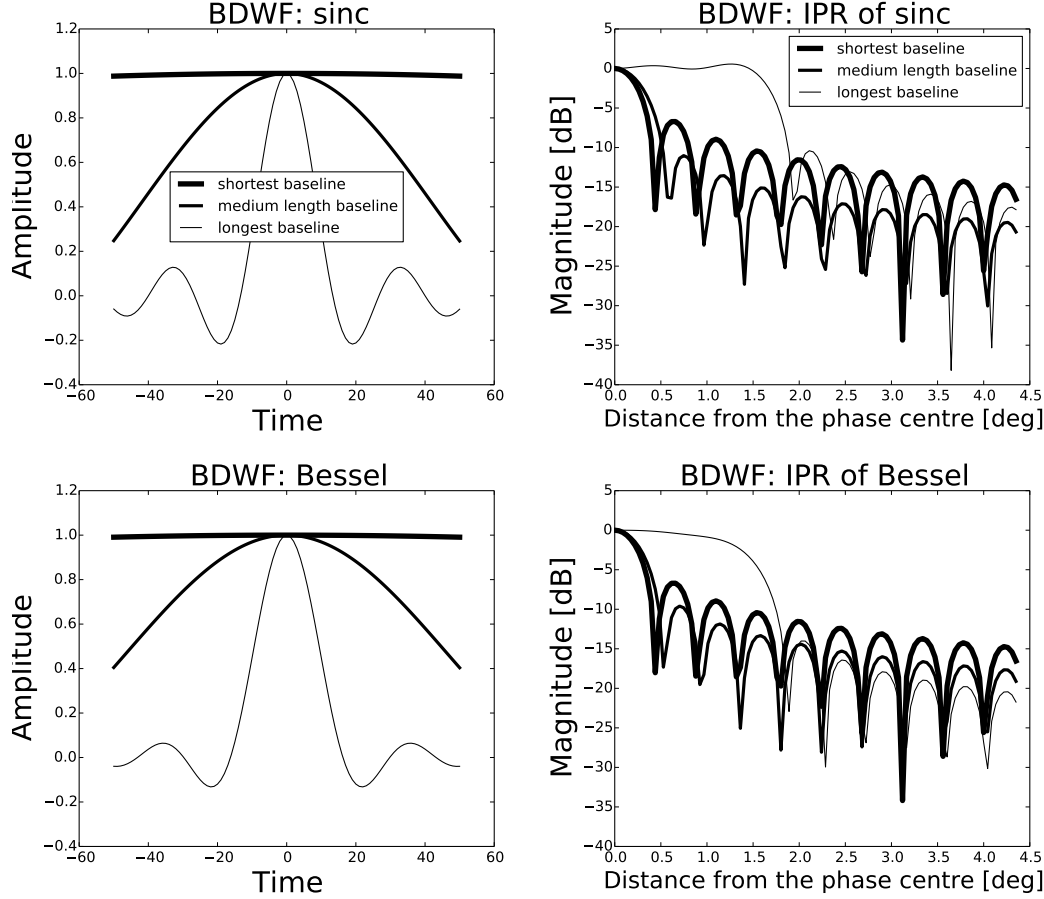
Here the FoI is adjustable by the parameter  $\kappa$  (in radians).

#### 3.1 Overlapping BDWFs

A more sophisticated approach involves overlapping BDWFs. Normal averaging implicitly assumes that the resampling bins  $\mathbf{B}_{kl}$  in eq. (36) do not overlap for adjacent  $kl$ , since they represent adjacent averaging intervals. There is, however, no reason (apart from computational load) why we cannot take the sum in eq. (36) over larger bins. Let us define the *window bin* for *overlap factors* of  $\alpha, \beta$  as

$$\mathbf{B}_{kl}^{[\alpha\beta]} = \{ij : t_i v_j \in \mathbf{B}_{kl}^{[\alpha\Delta t, \beta\Delta\nu]}\}, \quad (39)$$

i.e. as the set of sample indices corresponding to a bin of size  $\alpha\Delta t \times \beta\Delta\nu$  in  $tv$ -space. Let us then compute the sums in eq. (36) over the window bin. This becomes distinct from the resampling bin: while the latter represents the spacing of the resampled visibilities, the former represents the size of the window over which the convolution is computed. Only for  $\alpha = \beta = 1$  do the two bins become the same.



**Figure 4.** Cross-sections through two different BDWFs (top left: sinc, bottom left: Bessel) defined in  $uv$ -space, plotted along the time axis and their IPR (top right: IPR for sinc, bottom right: IPR for Bessel). This shows that the effective window function is a scaling and truncation of the underlying window function, with the shortest baselines reducing to a boxcar-like window function.

As mentioned above, the computational load for overlapping window functions will increase linearly as the factors  $\alpha$  and  $\beta$  increase. For a given resampling bin  $B_{kl}^{[\alpha\beta]}$  the computational complexity would scale as  $\mathcal{O}(\alpha\beta n_t n_\nu)$  compared to  $\mathcal{O}(n_t n_\nu)$ , which is the complexity for applying a non-overlapping window function to visibilities within the resampling bin  $B^{[\Delta t \Delta \nu]}$ . In this notation, the resampling bin for a non-overlapping window functions consists of  $n_t \times n_\nu$  samples, which is the number of multiplications between the baseline-dependent weights and the visibilities.

In the overlapping regime, the baseline-dependent weight for a single visibility is not defined by a unique BDWF, but by the strength of the correlation between the overall overlapping BDWFs with the visibility. BDWFs in the overlapping regime are schematically illustrated in Fig. 5. For simple averaging, overlapping offers no benefit, since it only broadens  $\Pi^{[uv]}$  and therefore increases smearing, but for a well-behaved BDWF, enlarging the window bin (while maintaining the same window function scale) means less truncation – thus lower sidelobes – and decreased noise, as more sampled visibilities are taken into account. On longer baselines, the IPR of a well-behaved overlapping BDWF means less smearing in the FoI and excellent out-of-FoI suppression (see Fig. B1). On shorter baselines, a well-behaved BDWF is equivalent to a boxcar and there-

fore enlarging the window size by overlapping results in a decrease of noise (see Sect. 3.2).

Appendix B provides an alternative way to look at the overlapping BDWFs applied to visibilities.

To distinguish overlapping BDWFs from non-overlapping ones, in the rest of the paper we will designate the window functions employed as  $WF\text{-}\alpha \times \beta$ . For example, sinc-3 $\times$ 2,  $J_0$ -1 $\times$ 1 (i.e. no overlap), etc. If resampling is only done in one direction (only time or only frequency), we will indicate this as e.g.  $J_0$ -3 $\times$ –.

### 3.2 Noise penalty estimates: analytic

Let us now work out analytically the *noise penalty* associated with replacing an unweighted average by a weighted sum. For simplicity, let us assume that the noise term has constant r.m.s.  $\sigma_s$  across all baselines and samples. If the resampling bin consists of  $n_{\text{avg}} = n_t \times n_\nu$  samples, and since the noise is not correlated between samples, the noise on the averaged visibilities in eq. (29) will be given by

$$\sigma_m^2 = \frac{1}{n_{\text{avg}}^2} \sum_{i=1}^{n_{\text{avg}}} \sigma_s^2 = \frac{\sigma_s^2}{n_{\text{avg}}} \quad (40)$$

Note that the noise is uncorrelated across averaged visibilities. We can therefore use the the imaging equation (16) to



derive the following expression for the variance of the noise term in each pixel of the dirty image:

$$\sigma_{\text{pix}}^2 = \frac{(\sum_{pqkl} W_{pqkl}^2 \sigma_m^2)}{(\sum_{pqkl} W_{pqkl})^2}, \quad (41)$$

which for natural image weighting ( $W \sim \sigma_m^{-1}$ , i.e.  $W \equiv 1$  in this case) is simply

$$\sigma_{\text{pix}}^2 = \frac{1}{N} \frac{\sigma_s^2}{n_{\text{avg}}}, \quad (42)$$

where  $N$  is the total number of visibilities used for the synthesis.

To simplify further notation, let us replace  $pqkl$  by a single index  $\mu$ , enumerating all the lo-res visibilities  $\hat{V}_\mu$ , with  $\mu = 1 \dots N$ . If we now employ eq. (36) to compute the lo-res visibilities using some BDWF  $X(u, v)$ , the noise term becomes different per each visibility  $\mu$ :

$$\sigma_{X\mu}^2 = \frac{\sum X^2(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl})}{[\sum X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl})]^2} \sigma_s^2 \quad (43)$$

where both sums are taken over the window bin,  $i, j \in \mathbf{B}_{kl}$ . Let us define the *visibility noise penalty* associated with BDWF  $X$  and visibility  $\mu$  as the relative increase in noise over the unweighted average, i.e.

$$\Xi_{X\mu} = \frac{\sigma_{X\mu}}{\sigma_m} = \frac{\sqrt{n_{\text{avg}} \sum X^2(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl})}}{\sum X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl})}. \quad (44)$$

Note that in the case of overlapping BDWFs, the window bin in eq. 43 is larger than the resampling bin, and contains  $n_X$  samples, with  $n_X = \alpha\beta n_{\text{avg}}$ , where  $\alpha$  and  $\beta$  are the overlap factors. For  $\alpha = \beta = 1$ , it is easy to see that  $\Xi_{X\mu} \geq 1$ , and only reaches 1 when  $X \equiv 1$ . In other words, non-overlapping BDWFs always result in a visibility noise penalty above 1, while overlapping BDWFs can actually *reduce* noise in the resampled visibilities.

While paradoxical at first glance, this reduction in noise does **not** result in a net gain in image sensitivity. The reason for this is that with overlap in effect, the noise terms become correlated across resampled visibilities  $kl$  (within the same baseline  $pq$ ), with each hi-res visibility sample contributing to multiple resampled visibilities, and the image noise term no longer follows eq. (41).

If the resampled visibilities correspond to a single-channel snapshot, or if the BDWFs are non-overlapping, then the noise across visibilities remains uncorrelated, and we can compute the *image noise penalty* associated with imaging weights  $W$  and BDWF  $X$  as

$$\Xi_X^W = \frac{\sigma_{\text{pix}, X}^2}{\sigma_{\text{pix}}^2} = \frac{n}{\sigma_s} \frac{(\sum_\mu W_\mu^2 \sigma_\mu^2)}{(\sum_\mu W_\mu)^2} = \frac{(\sum_\mu W_\mu^2 \Xi_{X\mu}^2)}{(\sum_\mu W_\mu)^2} \quad (45)$$

In the case of natural weighting ( $W_\mu = \sigma_\mu^{-1}$ ) this reduces to:

$$\Xi_X^{\text{nat}} = \frac{N}{\sum_\mu \Xi_{X\mu}^{-1}}. \quad (46)$$

### 3.3 Noise penalty estimates: empirical

In this section we employ simulations to empirically verify noise estimates computed using the derivation above. We generate a high resolution (“high-res”) VLA-C measurement

set (MS) corresponding to a 400 s synthesis with 1 s integration, with 30 MHz of bandwidth centred on 1.4 GHz, divided into 360 channels of 83.4 kHz each. The MS is filled with simulated thermal noise with  $\sigma_s = 1$  Jy. We then generate a low resolution (“low-resolution”) MS using 100 s integration, with a single frequency channel of 10 MHz. This MS receives the resampled visibilities. The size of the resampling bin is thus 100 s by 10 MHz, or  $100 \times 120$  in terms of the number of hi-res samples.

We then resample the hi-res visibilities using a number of BDWFs, and store the results in the lo-res MS:

- Standard averaging to 100 s and 10 MHz (using the middle 120 channels). This gives us the baseline noise estimate.
- sinc and Bessel windows using the same bin, without an overlap, tuned to a FoI of  $2^\circ$ .
- The same windows with overlap factors of  $4 \times 3$ .

We then image the lo-res MS and take the r.m.s. pixel noise across the image as an estimator of  $\sigma_{\text{pix}}$ , divide it by the baseline estimate produced with normal averaging, and compare the resulting noise penalty with that predicted by eq. 46. Note that this estimator is not perfect, since image noise is correlated across pixels. Nonetheless, we obtain results that are broadly consistent with analytical predictions (Table 1).

Fig. 6 shows the predicted visibility noise penalty for the same BDWFs, as a function of East-West baseline component, which determines the baseline rotation speed. Note that the noise penalty rises sharply towards longer East-West baselines. Note also that the penalty is well below 1 on shorter baselines, when overlapping is in effect. As mentioned in Sect. 3.1, this is because the BDWF becomes equivalent to boxcar averaging (i.e. unweighted averaging or simple averaging) and can only decrease the noise when overlap is in place. This is simple to understand analytically as follows: the noise term has constant r.m.s.  $\sigma_s$  across all baselines (including samples) and the resampling bin consists of  $n_t \times n_\nu$  samples. Since the noise is not correlated between samples, the noise on the averaged (unweighted averaging) visibilities in eq. (29) is given by:

$$\sigma_m^2 = \frac{\sigma_s^2}{n_t n_\nu}. \quad (47)$$

On shorter baselines, the BDWF  $X \equiv 1$  (i.e. unweighted averaging) and overlapping means enlarging the averaged resampling bin by a factor of  $\alpha\beta$  with  $\alpha > 1$  and/or  $\beta > 1$ . The noise on the averaged visibilities becomes:

$$\sigma_{X\mu}^2 = \frac{\sigma_s^2}{\alpha\beta n_t n_\nu}, \quad (48)$$

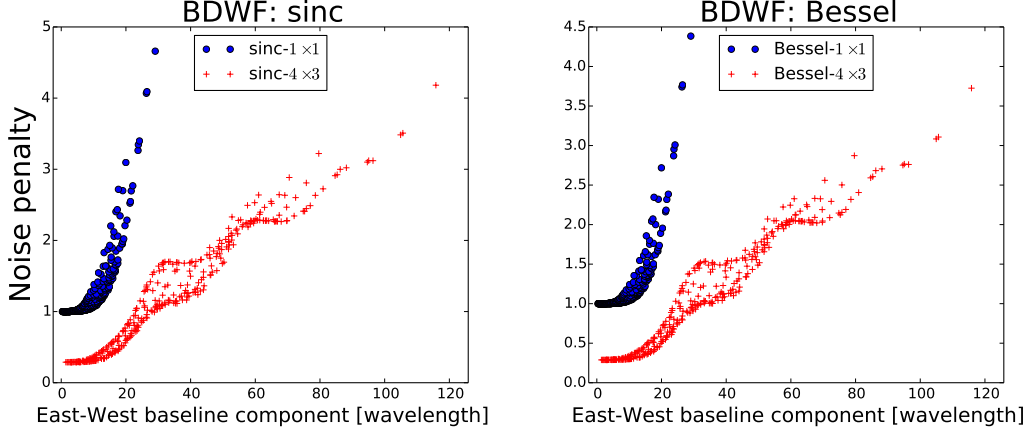
which gives us a noise penalty of

$$\Xi_{X\mu} = \frac{\sigma_{X\mu}}{\sigma_m} = \frac{1}{\sqrt{\alpha\beta}} < 1. \quad (49)$$

Note that as  $\alpha\beta > 1$ , it implies that  $\sqrt{\alpha\beta} > 1$ .

## 4 SIMULATIONS AND RESULTS

In this section we use BDWFs to resample simulated visibility data, and study the effect on smearing and source



**Figure 6.** Noise penalty w.r.t. normal averaging as a function of baseline length for two different BDWFs (sinc and Bessel), with and without overlap. VLA-C, 1.4 GHz, sampling intervals of 100 s and 10 MHz.

BDWF	$\Xi$ analytic	$\Xi$ sim
sinc-1 $\times$ 1	1.247	1.22
sinc-4 $\times$ 3	1.242	1.20
Bessel-1 $\times$ 1	1.178	1.16
Bessel-4 $\times$ 3	1.109	1.12

**Table 1.** A comparison of image noise penalties associated with different BDWFs, computed analytically vs. simulations.

suppression. Apart from a few examples documented separately, the basis interferometer configuration employed in the simulations corresponds to VLA-C observing at 1.4 GHz. Similarly to Sect. 3.3, we create a “high-res” measurement set corresponding to a 400 s synthesis at 1 s integration, with 30 MHz total bandwidth centred on 1.4 GHz, divided into 360 channels of 83.4 kHz each. The MS is populated by noise-free simulated visibilities corresponding to a single point source at a given distance from the phase centre. We then generate “low-res” MSs to receive the resampled visibilities, resample the latter using a range of BDWFs, convert the visibilities to dirty images (using natural weighting unless otherwise stated), and measure the peak source flux in each image. Since each dirty image corresponds to a single source, the peak flux gives us the degree of smearing or the smearing factor (i.e. the amplitude decrease for off-axis sources) associated with a given BDWF and sampling interval.

For the first set of simulations, the “low-res” MSs corresponds to a 100 s and 10 MHz synthesis. We employ three sampling rates, 25 s  $\times$  2.5 MHz, 50 s  $\times$  5 MHz and 100 s  $\times$  10 MHz (thus 4 timeslots and 4 channels, 2 timeslots and 2 channels, and single-channel snapshot).

A typical performance comparison for the VLA-C configuration at 1.4 GHz is given by Fig. 7. This figure illustrates some of the principal achievements of the present work, so let us spend some time explaining it. The horizontal axis represents distance from phase centre, while the vertical axis of the left-hand plot represents the smearing factor (left plot). Unity corresponds to no smearing; this is the case at phase centre, thus all curves start at unity. The three thick gray curves correspond to normal averaging into 25 s  $\times$  2.5 MHz, 50 s  $\times$  5 MHz and 100 s  $\times$  10 MHz. We can (rather arbitrarily) define a series of “acceptable” smearing

levels by specifying a FoI radius, and the maximum extent of smearing over that FoI. For the FoI radius, we may pick e.g. the half-power point of the primary beam, the main lobe of the primary beam, or extent of the first sidelobe of the primary beam. For VLA’s 25 m dishes at this frequency, these radii correspond to  $\sim 0.25^\circ$ ,  $0.5^\circ$ , and  $1^\circ$ , respectively; they are indicated by thin vertical lines in the figure. The thin horizontal line indicates our chosen smearing threshold of 0.95. In the right plot, all the curves are normalised with respect to the 25 s  $\times$  2.5 MHz averaging curve.

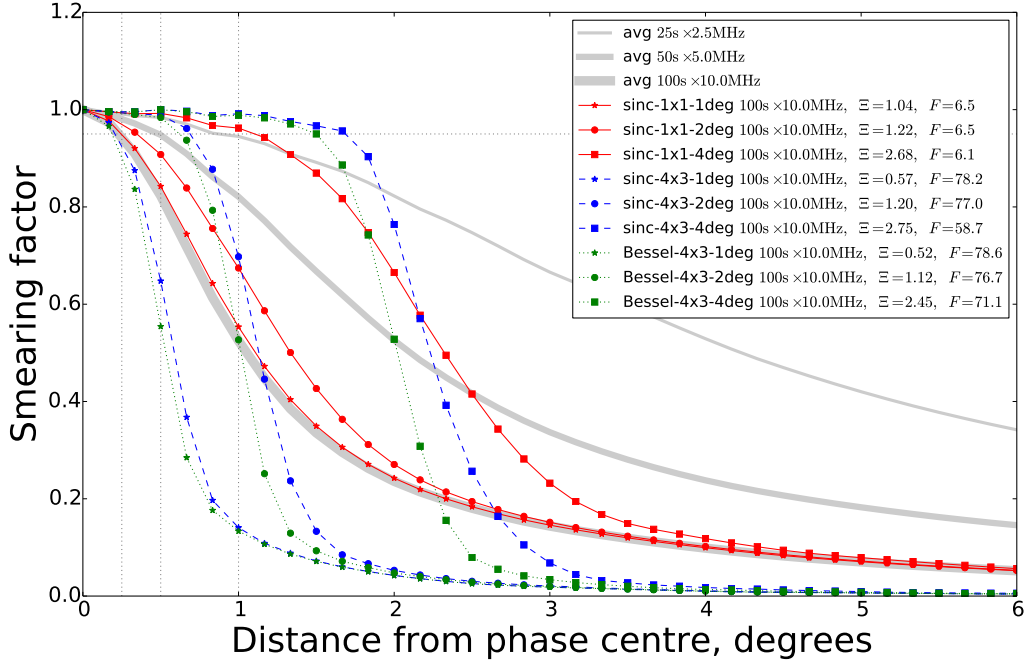
For regular averaging, the three chosen bin sizes happen to roughly correspond to acceptable levels of smearing over the three chosen FoI values. The other curves show the performance of a few different BDWFs, all at 100 s  $\times$  10 MHz sampling. There are three types of BDWFs shown, indicated by line style (and colour, in the colour version of the plot):

- sinc-1 $\times$ 1: a non-overlapping sinc window (solid line, red)
- sinc-4 $\times$ 3: an overlapping sinc window (dashed line, blue)
- Bessel-4 $\times$ 3: an overlapping Bessel window (dotted line, green)

These are tuned to three different FoI settings, as indicated by the plot symbol:  $1^\circ$  (star),  $2^\circ$  (circle),  $4^\circ$  (square).

The plot is meant to show performance of BDWFs at 100 s  $\times$  10 MHz versus a “baseline case” of 25 s  $\times$  2.5 MHz averaging, the latter being an acceptable averaging setting for this particular frequency and telescope geometry. The legend next to the plot therefore indicates  $\Xi$ , the noise penalty associated with that particular BDWF, and  $F$ , the far source suppression factor. Both values are calculated w.r.t. the baseline case. Note the following salient features:

- All overlapping BDWFs provide outstanding far source suppression in this regime, with  $F$  in the 60  $\sim$  80 range. The non-overlapping sinc (solid red lines) only achieves  $F$  6, which is similar to regular averaging at the same rate.
- Noise performance is excellent for the  $1^\circ$  BDWFs. There is a small noise penalty at  $2^\circ$ , and a larger (over a factor of 2) noise penalty at  $4^\circ$ . This can be easily understood by considering the shape of BDWFs as a function of FoI:



**Figure 7.** VLA-C observing at 1.4 GHz for the 400 s snapshot and 30 MHz band synthesis; illustrating the performance of non-overlapping and overlapping BDWFs tuned to three FoIs for the same  $uv$ -bin size and the associated noise penalty. Smearing as a function of distance from the phase centre, for conventional averaging with 25 s  $\times$  5 MHz and 100 s  $\times$  10 MHz bins, and for several BDWFs with 100 s  $\times$  10 MHz bins. The noise penalty  $\Xi$  and the far-source suppression factor  $F$  are given relative to 25 s  $\times$  2.5 MHz averaging.

smaller FoIs correspond to broader windows that become more “boxcar-like” over the sampling interval, and vice versa. This means that, in this particular configuration, BDWFs cannot achieve a FoI of  $4^\circ$  at 10 s  $\times$  100 MHz without a substantial sacrifice in sensitivity. We shall return to this issue below.

- If the desired FoI size is  $r \sim 0.5 - 1^\circ$ , overlapping BDWFs (sinc-4x3-2deg and Bessel-4x3-2deg) provide excellent performance at 100 s  $\times$  10 MHz. Compared to averaging at 25 s  $\times$  2.5 MHz, they achieve a factor of 16 data compression with minimal loss of sensitivity, with excellent tapering behaviour: the smearing performance across the FoI is equivalent to (or better) than that of normal averaging, and out-of-FoI source suppression is almost two orders of magnitude higher.

Fig. 8 presents the same results in an alternative way. Here, the recovered flux is shown relative to the baseline case of 25 s  $\times$  2.5 MHz averaging. This clearly illustrates the excellent performance of overlapping BDWFs tuned to  $2^\circ$ .

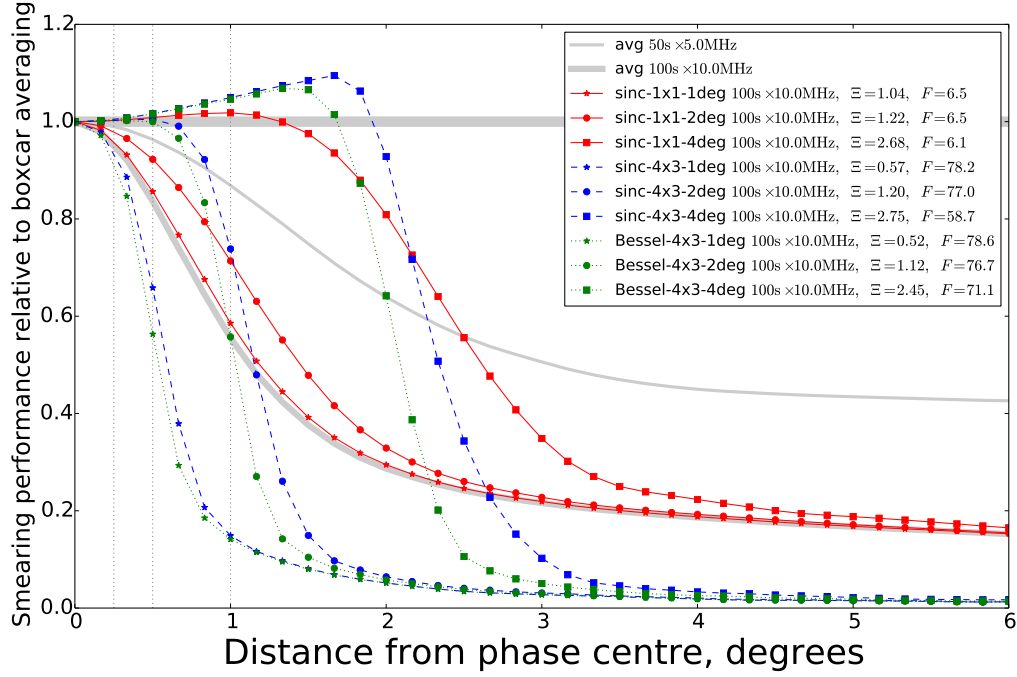
#### 4.1 Noise penalties and overlapping BDWFs

Values of  $\Xi < 1$  above may be paradoxical at first, since one can not theoretically exceed the noise performance of the unweighted average. This, however, is an artefact of our short simulation. Overlapping BDWFs are essentially averaging in “bonus signal” from regions of overlap extending outside the

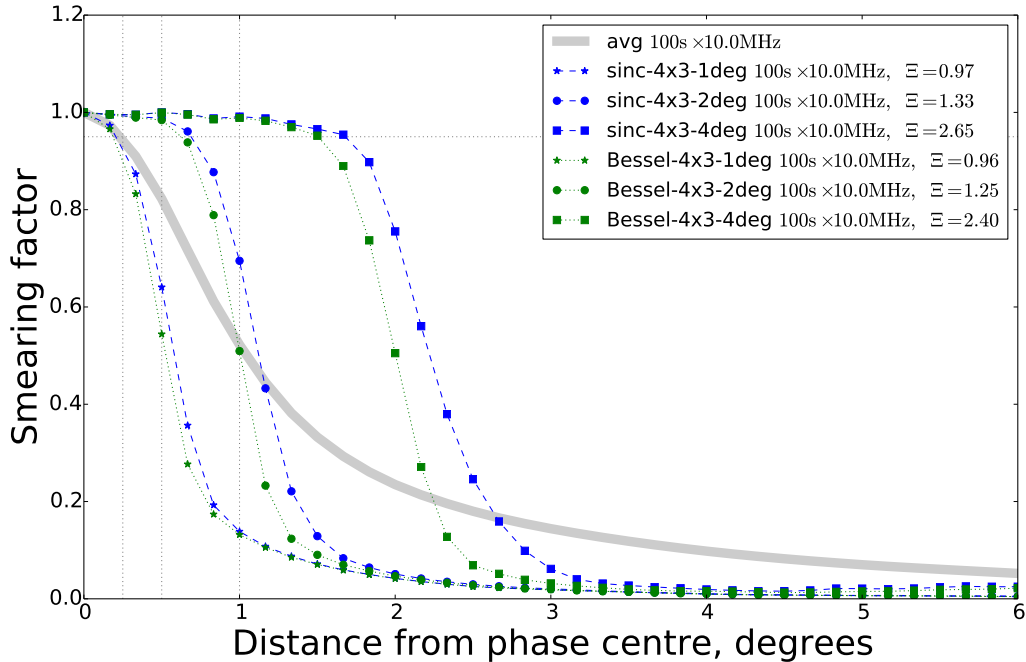
nominal time and frequency coverage. In our case, at 100 s  $\times$  10 MHz sampling, a BDWF with  $4 \times 3$  overlap is actually adding up signal over a 400 s  $\times$  30 MHz bin, i.e. a bin that is a factor of 12 larger (though of course the bonus sensitivity thus gained is much less than the theoretically available  $\sqrt{12}$ , since the weights over the overlap regions correspond to the “wings” of the window function, and are thus small). This can easily result in lower per-visibility noise than that achieved by regular averaging over 100 s  $\times$  10 MHz, and correspondingly higher snapshot sensitivity.

In the more realistic case of a long, multiple-channel synthesis (what we will call a *full synthesis*), the effects of bonus sensitivity disappear. While the noise on individual visibilities remains nominally lower in a full synthesis thanks to the overlap, it becomes correlated across neighbouring  $uv$ -bins, so there is no net gain in image-plane sensitivity. Strictly speaking, at the “edge” of the synthesis, overlapping BDWFs are still pulling in some bonus signal from overlap regions extending beyond the synthesis coverage, but since the area of this overlap is negligible compared to the coverage of the full synthesis, so is the effect of the bonus signal.

In other words, simulating a snapshot observation results in underestimated noise penalties, compared to the real-life case of a full synthesis. We should expect the noise penalties to go up (and eventually exceed unity) as we increase the synthesis time and number of channels. Fig. 9 presents the results of such a simulation. This shows a 1800 s  $\times$  200 MHz synthesis, sampled at the same rates as above. The results should be compared to and contrasted



**Figure 8.** VLA-C observing at 1.4 GHz. Results of Fig. 7 normalised to the  $25 \text{ s} \times 2.5 \text{ MHz}$  averaging curve, the horizontal grey line at  $y=1$  indicates that the  $25 \text{ s} \times 2.5 \text{ MHz}$  averaging curve is normalised by itself. This figure clearly illustrates the excellent performance of overlapping BDWFs tuned to  $2^\circ$ .



**Figure 9.** VLA-C observing at 1.4 GHz for the full 30 min and 200 MHz band synthesis; illustrating the performance of overlapping BDWFs tuned to three FoIs for the same  $uv$ -bin size, and the associated noise penalty. Smearing as a function of distance from the phase centre, for conventional averaging with  $100 \text{ s} \times 10 \text{ MHz}$  bins, and for overlapping BDWFs with  $100 \text{ s} \times 10 \text{ MHz}$  bins. The noise penalty  $\Xi$  is given relative to  $100 \text{ s} \times 10 \text{ MHz}$  averaging.

with those of Fig. 7. Note that the tapering response of BDWFs is nearly identical, while the noise penalties are indeed higher. With  $4 \times 3$  overlap and  $100 \text{ s} \times 10 \text{ MHz}$  sampling, the total signal accessed by overlapping BDWFs corresponds to  $2100 \text{ s} \times 220 \text{ MHz}$ , which gives a theoretical maximum of a factor of  $\sim 1.13$  in bonus sensitivity. In other words, the values of  $\Xi$  in Fig. 9 are still underestimated, but by 13% at most (which explains  $\Xi < 1$  for the  $1^\circ$  case). From this we may safely extrapolate that the noise penalty of BDWFs matched to  $1 - 2^\circ$  FoIs will remain reasonable even for a much longer and wider band synthesis.

#### 4.2 FoIs and sampling rates

For BDWFs, a given FoI tuning represents a characteristic scale in the  $uv$ -plane, which is inversely proportional to the FoI parameter. On the other hand, the  $uv$ -bin sampled by any given visibility is proportional to the integration time, fractional bandwidth, and baseline length. Since the window function is truncated at edge of the averaging bin (which can be larger by the sampling bin by a factor of several, if overlapping BDWFs are employed), there is, for any given baseline, some kind of optimal range of  $uv$ -bin sizes over which BDWFs tuned to a particular FoI setting are “efficient”. Over smaller bins, BDWFs become equivalent to a boxcar averaging, over larger bins, BDWFs penalise too much sensitivity as they down-weight more samples. Since this optimal bin size is proportional to baseline length, the overall optimum is dependent on the distribution of baselines in the array.

Furthermore, the sampling rate needs to be “balanced” in time and frequency for BDWFs to achieve efficient tapering response. If the  $uv$ -bins are elongated, the window function becomes truncated (i.e. more boxcar-like) across the bin, which reduces its ability to induce the desired taper. Since the orientation of the bins changes as the baseline rotates, the cumulative effect is an average degradation of the tapering response, in the sense that it becomes closer to that of boxcar averaging. In this sense, the optimal  $uv$ -bin shape is square-like. This occurs when the fractional bandwidth is equal to the arc section swept out by the baseline over one bin. For a polar observation (circular  $uv$ -tracks), we can express this as

$$\Delta\nu/\nu = 2\pi(B_x/B)(\Delta t/24\text{h}), \quad (50)$$

where  $B$  is the baseline length, and  $B_x$  is its East-West component. Rewriting this in terms of more convenient units, we obtain

$$\frac{\Delta\nu_{\text{MHz}}}{\Delta t_{\text{s}}} \approx \frac{\nu_{\text{MHz}}}{14000} \cdot \frac{B_x}{B}, \quad (51)$$

leading to a simple rule-of-thumb: at 1.4 GHz, an East-West baseline sweeps out a square-like bin when the integration time in seconds is 10 times the channel width in MHz (hence the use of bin sizes such as  $100 \text{ s} \times 10 \text{ MHz}$  in the analysis here).

The interaction between  $uv$ -bin size and tapering response is illustrated in Fig. 10. Here we compare the performance of two BDWFs tuned to a  $4^\circ$  FoI – a non-overlapping sinc- $1 \times 1$  filter (solid red lines) and an overlapping sinc- $4 \times 3$  filter (dashed blue lines) – over three sampling bin sizes:  $25 \text{ s} \times 2.5 \text{ MHz}$ ,  $50 \text{ s} \times 5 \text{ MHz}$  and  $100 \text{ s} \times 10 \text{ MHz}$ . For reference, the performance of boxcar averaging over the same

bin sizes is indicated by the thick grey lines. Note how at the smaller bin size, the non-overlapping sinc is practically equivalent to a boxcar in terms of tapering response; at the larger bin size, it begins to shape the FoI. Introducing an overlap improves the response considerably. An overlapping filter at  $25 \text{ s} \times 2.5 \text{ MHz}$  achieves almost the same tapering response as a non-overlapping one at  $100 \text{ s} \times 10 \text{ MHz}$  (which is not surprising, if one considers that the effective averaging bin size in the former case is  $100 \text{ s} \times 7.5 \text{ MHz}$ ). However, for all filters, at  $100 \text{ s} \times 10 \text{ MHz}$  the noise penalty goes up quite sharply.

This illustrates that  $50 \text{ s} \times 5 \text{ MHz}$  is an appropriate BDWF sampling rate for achieving a  $4^\circ$  FoI (for VLA-C configuration at 1.4 GHz), providing a reasonable trade-off between tapering response and noise penalty. At higher sampling rates, the tapering response is degraded, while at lower sampling rates, the noise penalty increases. In comparison (as we saw in the previous section), for FoIs of  $1 - 2^\circ$ , BDWFs achieve a good trade-off at  $100 \text{ s} \times 10 \text{ MHz}$  sampling.

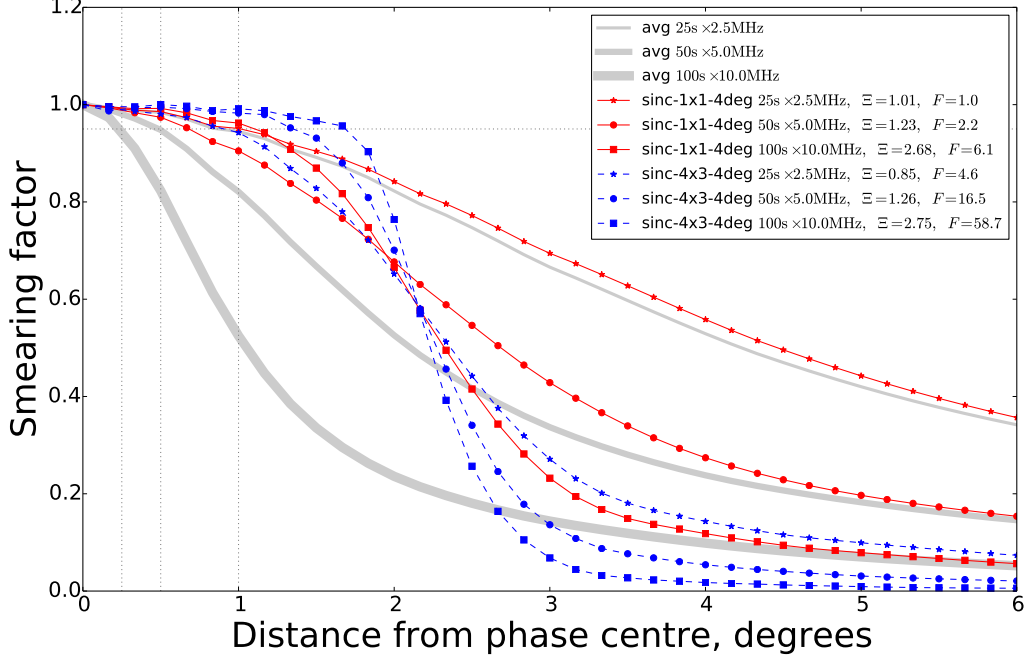
It is interesting to consider how optimal BDWF sampling changes as a function of array size. Fig. 11 shows a simulation for VLA-C at 14 GHz. (Since our results are completely determined by  $uv$ -plane geometry in wavelengths, this is equivalent to VLA-C scaled up by a factor of 10 at an observing frequency of 1.4 GHz). From eq. 51, we can see that square-like  $uv$ -bins correspond to sampling rate combinations such as  $10 \text{ s} \times 10 \text{ MHz}$ . The simulation presented here is for a  $1800 \text{ s} \times 200 \text{ MHz}$  synthesis, i.e. is closer to the full synthesis rather than a snapshot case. Comparing Figs. 11 and 9, we find nearly identical BDWF performance (in terms of tapering response and noise penalty) at 14 GHz and 1.4 GHz, with only the optimal sampling rate being different.

#### 4.3 BDWFs for wide-field VLBI

In the VLBI case, it is usually a combination of smearing and data rates, rather than the primary beam, that effectively limits the FoI. For example, the current European VLBI Network (EVN) correlator (Keimpema et al. 2015) operated by the Joint Institute for VLBI in Europe (JIVE) is capable of producing data at dump rates down to 10 ms, with 16 MHz of total bandwidth split into up to 8192 channels. The maximum available FoI of an EVN experiment is restricted by the smallest primary beam, which is usually that of the 100 m Effelsberg telescope – or about  $10'$  in diameter at L-band. The ENV calculator<sup>3</sup> shows that a dump rate of 0.125 s and 1024 channels (16 kHz) is required to keep smearing to within 10% across this FoI. Due to the large computational and storage requirements, such data rates have only been employed in one-off experiments. For routine use, techniques such as multiple-phase centre correlation are more common. Typically, data is averaged into more modest sampling rates of 2 s and 32 channels. This restricts the effective (L-band) FoI to about  $20''$ , and thus limits the scientific usefulness of archival data to narrow-FoI experiments.

In this section we investigate whether the use of BDWFs can enable true wide-field VLBI. We simulate a 1.6 GHz EVN observation employing eight stations (Effelsberg,

<sup>3</sup> <http://www.evlbi.org/cgi-bin/EVNcalc>



**Figure 10.** VLA-C observing at 1.4 GHz for the 400 s snapshot and 30 MHz band synthesis; illustrating the performances of non-overlapping and overlapping BDWFs tuned to  $4^\circ$  FoI and for three  $uv$ -bin size, with the associated noise penalty. Smearing as a function of distance from the phase centre, for conventional averaging with  $25\text{ s} \times 2.5\text{ MHz}$ ,  $50\text{ s} \times 5\text{ MHz}$  and  $100\text{ s} \times 10\text{ MHz}$  bins, and for several BDWFs with  $100\text{ s} \times 10\text{ MHz}$  bins. The noise penalty  $\Xi$  and the far-source suppression factor  $F$  are given relative to  $25\text{ s} \times 2.5\text{ MHz}$  averaging.

Hartebeesthoek, Jodrell Bank, Noto, Onsala, Torun, VLA, Westerbork, Shanghai), with a maximum baseline of 10161 km. Fig. 12 compares the smearing response of normal averaging to that of two overlapping Bessel BDWFs, employing 0.5 s and 25 kHz sampling. At these data rates, it becomes almost practical to have a full-FoI EVN archive.

For comparison, we also show the performance of BDWFs for a hypothetical fast-transient archive application. In order to localise potential fast radio bursts (FRBs), we would need to retain a high time resolution of 10 ms, as well as a lower spectral resolution for de-dispersion. In this regime, BDWFs are less efficient since the  $uv$ -bins are elongated. Fig. 12 shows that this translates into less source suppression outside the FoI, but does not impact the ability to retain FoI.

## 5 CONCLUSIONS AND FUTURE WORK

The goal of this work was to demonstrate the application of baseline-dependent window functions to radio interferometry. We have demonstrated that BDWFs offer a number of interesting advantages over conventional averaging. The first of these is data compression – i.e. visibilities can be sampled at a lower rate, while retaining a large FoI. Compression by a factor of 16 with relatively little loss of sensitivity has been demonstrated. The huge data rates from upcoming instruments such as ASKAP, MeerKAT and the future SKA1 mean that raw visibilities may need to be discarded after

calibration (unlike older instruments, where raw visibilities have typically been archived). This represents something of a risk to the science, as it precludes future improvements in calibration techniques from being later applied to the data. With BDWFs, at least a highly-compressed version of the visibilities may be retained.

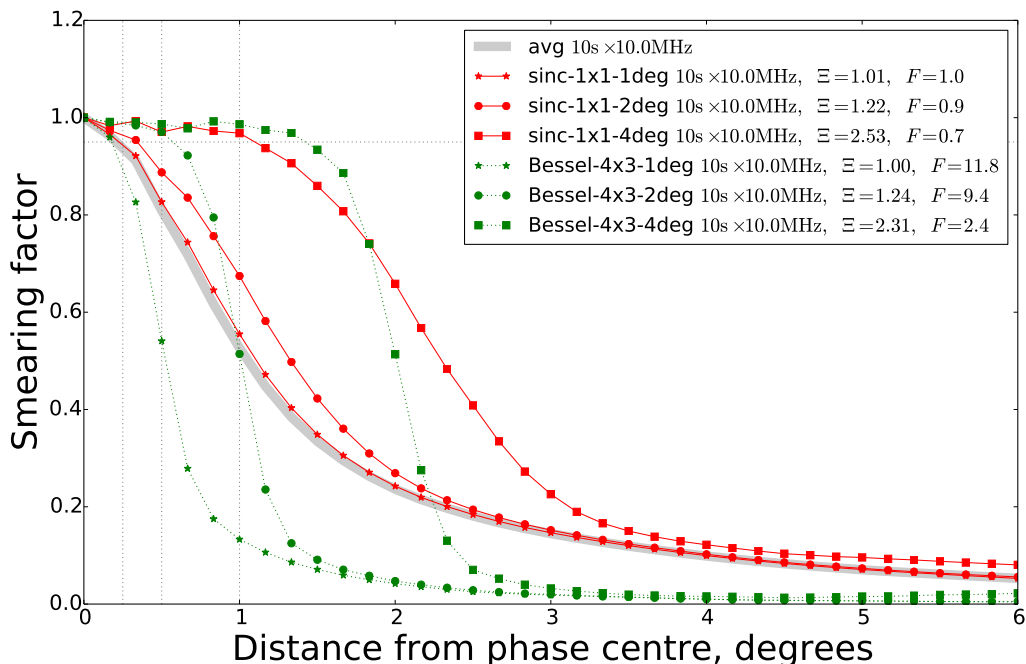
The second potential benefit of BDWFs is the increased suppression of unwanted signal from out-of-FoI sources. This reduces both the overall level of far sidelobe confusion noise, and lessens the impact of A-team sources in sidelobes.

Thirdly, BDWFs can have an interesting impact in the VLBI case, as they allow the full primary beam FoI to be imaged using a single VLBI dataset. This opens the door to wide-field VLBI, which has previously been impractical.

BDWFs have a number of potential downsides. The first one is a potential loss in sensitivity. Our simulations show that this can be kept within reasonable limits, especially if overlapping BDWFs are employed, and can be traded off with compression rate. Note that from DSP theory, we know that a properly matched filter can actually increase the SNR: the equivalent result with BDWFs is that off-axis sources are attenuated less, so the effective SNR off-axis can increase despite the loss in absolute sensitivity.

The second downside of BDWFs is an increase in computational complexity. Whether implemented in a correlator or in post-processing, BDWFs (and especially overlapping BDWFs) require substantially more operations than simple averaging. There may be other limits to the practical applicability of BDWFs. They are far less efficient if high spectral





**Figure 11.** VLA-C observing at 14 GHz. This shows that the results of the BDWFs are completely determined by the  $uv$ -plane geometry in wavelengths. This figure is equivalent to VLA-C observing at a frequency of 1.4 GHz, with baselines scaled up by a factor of 10. Smearing as a function of distance from the phase centre, for conventional averaging with  $2.5 \text{ s} \times 2.5 \text{ MHz}$ ,  $5 \text{ s} \times 5 \text{ MHz}$  and  $10 \text{ s} \times 10 \text{ MHz}$  bins, and for several BDWFs with  $10 \text{ s} \times 10 \text{ MHz}$  bins. The noise penalty  $\Xi$  and the far-source suppression factor  $F$  are given relative to  $2.5 \text{ s} \times 2.5 \text{ MHz}$  averaging.

resolution is required, so their use may be limited to continuum observations. Furthermore, averaging over longer intervals requires accurate phase calibration, so high compression rates may only be achievable post-calibration.

An interesting avenue of future research is combining BDWFs with baseline-dependent averaging. As we saw above, the ability of BDWFs to shape a FoI is somewhat limited by the fact that shorter baselines sweep out smaller bins in  $uv$ -space, with window functions over them becoming boxcar-like. If baseline-dependent averaging is employed, shorter baselines are averaged over larger  $uv$ -bins, thus increasing the effect of BDWFs.

Finally, we should note that the use of BDWFs results in a different position-dependent PSF than regular averaging (or to put it another way, the smearing response of BDWFs results in a different smeared PSF shape). Future work will focus on methods of deriving this PSF shape, with a view to incorporating this into current imaging algorithms.

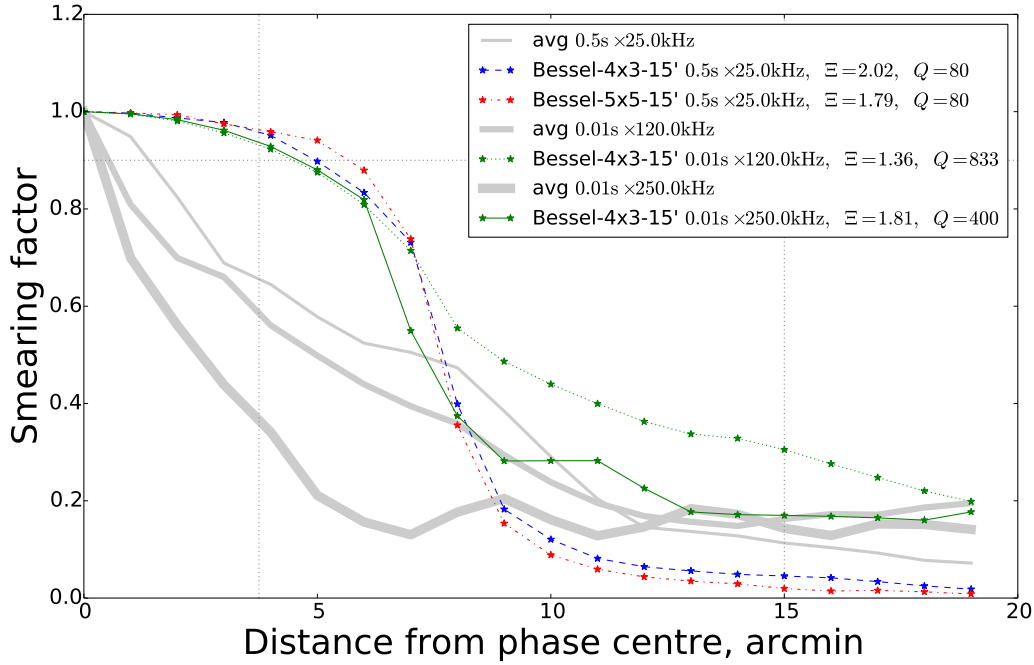
## ACKNOWLEDGEMENTS

This work is based upon research supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation. The EVN-related research in this paper emerged from fruitful discussions with Dr Aard Keimpema and Dr Zsolt Paragi at the Joint Institute for VLBI ERIC, whom we would like to thank for this collaboration. The visit to JIVE ERIC was

made possible by the FP7 MIDPREP program. We would like to thank the anonymous referee for comments that substantially improved the paper.

## REFERENCES

- Bregman J. D., 2012, PhD thesis, Ph. D. Thesis, University of Groningen, Groningen, The Netherlands, pp81-82
- Bridle A., Schwab F., 1999, in *Synthesis Imaging in Radio Astronomy II* Vol. 180, Bandwidth and time-average smearing. p. 371
- Bridle A. H., Schwab F. R., 1989, in *Synthesis Imaging in Radio Astronomy* Vol. 6, Wide field imaging i: Bandwidth and time-average smearing. p. 247
- Keimpema A., Kettenis M., Pogrebenko S., Campbell R., Cimó G., Duev D., Eldering B., Kruithof N., van Langevelde H., Marchal D., et al., 2015, *Experimental Astronomy*, pp 1–21
- Noordam J. E., Smirnov O. M., 2010, *A&A*, 524, A61
- Offringa A. R., de Bruyn A. G., Zaroubi S., 2012, *MNRAS*, 422, 563
- Perley R., 2013, High Dynamic Range Imaging, presentation at “The Radio Universe @ Ger’s (wave)-length” conference (Groningen, November 2013), <http://www.astron.nl/gerfeest/presentations/perley.pdf>
- Prabhu K., 2013, Window functions and their applications in signal processing. CRC Press



**Figure 12.** VLBI simulation (1016 1km maximum baseline) at 1.6 GHz. Smearing as a function of distance from the phase centre, for conventional averaging at  $0.5 \text{ s} \times 25 \text{ kHz}$ ,  $10 \text{ ms} \times 120 \text{ kHz}$  and  $10 \text{ ms} \times 250 \text{ kHz}$  averaging, and overlapping Bessel BDWFs. The horizontal line corresponds to 10% smearing losses. The ratio  $Q$  shows the increase in data volume over conventional  $2 \text{ s} \times 0.5 \text{ MHz}$  averaging.

- Smirnov O. M., 2011, *Astronomy & Astrophysics*, 527, A106
- Smirnov O. M., Frank B., Theron I. P., Heywood I., 2012, in *Int. Conf. on Electromagnetics in Advanced Applications (ICEAA 2012)*. 2-7 September. Cape Town, South Africa Understanding the impact of beamshapes on radio interferometer imaging performance
- Smith S. W., et al., 1997
- Thompson A. R., 1999, in *Synthesis Imaging in Radio Astronomy II Vol. 180, Fundamentals of radio interferometry*. p. 11
- Thompson A. R., Moran J. M., Swenson, Jr. G. W., 2001, *Interferometry and Synthesis in Radio Astronomy*, 2 edn. Wiley, New York
- Thompson A. R., Moran J. M., Swenson Jr G. W., 2008, *Interferometry and synthesis in radio astronomy*. John Wiley & Sons
- van Haarlem M., Wise M., Gunst A., Heald G., McKean J., Hessels J., De Bruyn A., Nijboer R., Swinbank J., Fallows R., et al., 2013, *Astronomy & astrophysics*, 556, A2
- Watson G. N., 1995, *A treatise on the theory of Bessel functions*. Cambridge university press
- Wrobel J., Walker R., 1999, in *Synthesis Imaging in Radio Astronomy II Vol. 180, Sensitivity*. p. 171

## APPENDIX A: RELATIVE PERFORMANCE OF WINDOW FUNCTIONS

Window functions – or rather their corresponding image-plane response (IPR) – can be characterised in terms of various metrics. Some common ones are the peak sidelobe level (PSL), the main lobe width (MLW) and the sidelobes roll-off (SLR) rate. In terms of the “ideal” IPR (Fig. 2, left), these correspond to the following desirable traits:

- Maximally conserve the signal within the FoI (Regime 1 in the figure), and make the transition region (Regime 2) as sharp as possible. Both of these correspond to larger main lobe width.
- Attenuate sources outside the FoI (Regime 3): this corresponds to a lower peak sidelobe level and higher sidelobes roll-off.

Table A1 summarises the performance of the different window functions. This table shows that the sinc, the Bessel and all their derivative with Hamming, Han and Blackman window functions have large main lobe, low peak sidelobe level and high sidelobes roll-off and therefore provide the more optimal tapering response.

## APPENDIX B: OVERLAPPING BDWFs

Let us reconsider that  $X$  is a BDWF with its appropriate resampling bin  $B_{kl}^{[\Delta t \Delta \nu]}$ . Now, suppose that  $\Delta^{\text{ift}} t$  and  $\Delta^{\text{ift}} \nu$  are the overlap time-frequency sampling intervals associated



window functions		~MLW (deg and at -3dB)	PSL (dB)	SLR (dB/oct)
$\Pi(t/t_a)$	$t \in  t_a $	1.406	-6.663	-12.089
	$t \in  t_a/2 $	2.812	-6.671	-11.065
$\text{sinc}(t)$	$t \in  t_a $	12.304	-10.889	-12.661
	$t \in  t_a/2 $	12.304	-13.241	-11.447
$J_0(t)$	$t \in  t_a $	9.140	-14.553	-12.011
	$t \in  t_a/2 $	9.140	-13.614	-11.794
$G(t)$	$b=3$	2.109	-21.535	-9.589
	$b=5$	2.812	-30.211	-9.091
$BW(t)$	$p=1$	2.109	-13.718	-12.581
	$p=3$	4.218	-10.145	-27.330
$PS_0(t)$	$\alpha = 2\pi^3$	3.515	-45.302	-7.424
	$\alpha = 5\pi^3$	4.218	-73.597	-6.375
$J_0^{\text{Hm}}(t)$	$t \in  t_a $	9.140	-35.724	-11.948
	$t \in  t_a/2 $	9.140	-22.670	-19.527
$\text{sinc}^{\text{Hm}}(t)$	$t \in  t_a $	12.656	-27.581	-13.817
	$t \in  t_a/2 $	12.656	-13.469	-14.324

**Table A1.** Comparative performance of different window functions. MLW: main lobe width, PSL: peak sidelobe level, SLR: sidelobes roll-off. Here  $\Pi$ : boxcar window,  $J_0$ : Bessel function of the first kind of order 0 (Watson 1995),  $G$ : Gaussian window,  $BW$ : Butterworth window,  $PS_0$ : prolate spheroidal wave function of sequence zero,  $\text{sinc}^{\text{Hm}}(t)$ : sinc multiplied by the Hamming function,  $J_0^{\text{Hm}}(t)$ : Bessel multiplied by the Hamming function.

with  $\Delta t$  and  $\Delta \nu$  on their *left-hand side*. Similarly,  $\Delta^{\text{rgt}} t$  and  $\Delta^{\text{rgt}} \nu$  are the overlap time-frequency sampling intervals associated with  $\Delta t$  and  $\Delta \nu$  on their *right-hand side*. The overlapping BDWFs sampling intervals are then given by:

$$\Delta^{\text{olp}} t = \Delta^{\text{lft}} t \cup \Delta t \cup \Delta^{\text{rgt}} t \quad (\text{B1})$$

$$\Delta^{\text{olp}} \nu = \Delta^{\text{lft}} \nu \cup \Delta \nu \cup \Delta^{\text{rgt}} \nu. \quad (\text{B2})$$

The overlap sampling bin is now defined as:

$$\begin{aligned} B_{kl}^{[\Delta^{\text{olp}} t, \Delta^{\text{olp}} \nu]} &= \left[ t_k - \frac{\Delta^{\text{olp}} t}{2}, t_k + \frac{\Delta^{\text{olp}} t}{2} \right] \\ &\times \left[ \nu_l - \frac{\Delta^{\text{olp}} \nu}{2}, \nu_l + \frac{\Delta^{\text{olp}} \nu}{2} \right] \end{aligned} \quad (\text{B3})$$

$$= B_{kl}^{\text{lft}} \cup B_{lk}^{[\Delta t \Delta \nu]} \cup B_{kl}^{\text{rgt}}, \quad (\text{B4})$$

where  $B_{kl}^{\text{lft}}$  and  $B_{kl}^{\text{rgt}}$  are the set of time-frequency samples in the overlap regimes, and  $B_{kl}^{[\Delta t \Delta \nu]}$  is the resampling bin for a non-overlapping BDWFs defined in eq. (20). Fig. B1 (top) displays the time direction overlap regimes (i.e.  $B_{kl}^{\text{lft}}$  and  $B_{kl}^{\text{rgt}}$ ). We note that the window bin defined in eq. (39) for overlap factors of  $\alpha, \beta$  is equivalent to

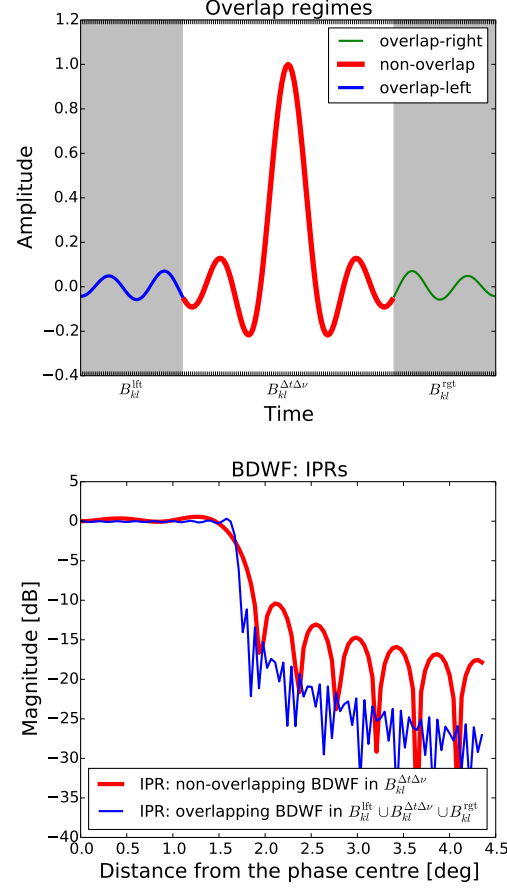
$$B_{kl}^{[\alpha\beta]} = \{ij : t_i \nu_j \in B_{kl}^{[\Delta^{\text{olp}} t, \Delta^{\text{olp}} \nu]}\} \quad (\text{B5})$$

or in more detail we can write:

$$\begin{aligned} B_{kl}^{[\alpha\beta]} &= \{ij : t_i \nu_j \in B_{kl}^{\text{lft}}\} \cup \{ij : t_i \nu_j \in B_{lk}^{[\Delta t \Delta \nu]}\} \\ &\cup \{ij : t_i \nu_j \in B_{kl}^{\text{rgt}}\} \end{aligned} \quad (\text{B6})$$

$$= S_1 \cup B_{kl} \cup S_2, \quad (\text{B7})$$

where  $S_1 = \{ij : t_i \nu_j \in B_{kl}^{\text{lft}}\}$  and  $S_2 = \{ij : t_i \nu_j \in B_{kl}^{\text{rgt}}\}$  are the number of time-frequency samples in the left-hand and right-hand overlap regimes respectively. The resampled visibilities for overlapping BDWFs (compare to eq. (36)) are



**Figure B1.** Overlapping BDWF. (Top) Blue curve: left-hand overlap function, green curve: right-hand overlap function, red curve: non-overlap function. (Bottom) The IPRs of an overlapping (blue) and non-overlapping (red) BDFW.

as follows:

$$\hat{V}_{pqkl} = \frac{\sum_{i,j \in B_{kl}^{[\alpha\beta]}} V_{pqij}^S X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl})}{\sum_{i,j \in B_{kl}^{[\alpha\beta]}} X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl})}. \quad (\text{B8})$$

One may still decompose eq. (B8) as:

$$\hat{V}_{pqkl} = \frac{Q^{\text{lft}} + \sum_{i,j \in B_{kl}} V_{pqij}^S X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl}) + Q^{\text{rgt}}}{R^{\text{lft}} + \sum_{i,j \in B_{kl}} X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl}) + R^{\text{rgt}}}, \quad (\text{B9})$$

where:

$$Q^{\text{lft}} = \sum_{i,j \in S_1} V_{pqij}^S X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl}),$$

$$Q^{\text{rgt}} = \sum_{i,j \in S_2} V_{pqij}^S X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl}),$$

$$R^{\text{lft}} = \sum_{i,j \in S_1} X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl}),$$

$$R^{\text{rgt}} = \sum_{i,j \in S_2} X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl}).$$

Fig. B1 shows an overlapping BDWF  $X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl})$  with  $ij \in S_1$  (or  $t_i \nu_i \in B_{kl}^{\text{lft}}$ ),  $X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl})$  with  $ij \in B_{kl}$

(or  $t_i \nu_i \in \mathbf{B}_{lk}^{[\Delta t \Delta \nu]}$ ) and  $X(\mathbf{u}_{pqij} - \mathbf{u}_{pqkl})$  with  $ij \in S_2$  (or  $t_i \nu_j \in \mathbf{B}_{lk}^{\text{rgt}}$ ). We note that eq. (B9) becomes equivalent to the non-overlapping resampled visibilities in eq. (36) when  $S_1 = S_2 = \emptyset$ . This is due to the sum over an empty set  $\emptyset$ :  $Q^{\text{lt}} = Q^{\text{rgt}} = R^{\text{lt}} = R^{\text{rgt}} = 0$ .